Journal of Applied Sciences Research, 6(2): 107-120, 2010
© 2010, INSInet Publication

Comprison of Model Selection Critaria for Multivariate Regression Model with Mixed Model

Ali Hussein AL-Marshadi

Department of Statistics, King Abdul Aziz University, Jeddah, Saudi Arabia

Abstract: This article considers the analysis of multivariate regression experiment that is used frequently in variety of applications research. We used simulation study to compare five model selection criteria in terms of their ability to identify the right multivariate regression model that has the right covariance structure and in the same time the right multivariate model structure. The comparison of the five model selection criteria was in terms of their percentage of number of times that they identify the right model. The simulation results indicate that overall, the percentages of identifying the right multivariate regression model from both standard and non-standard multivariate model structures were low except for specific models that involve the indicator variable. In the same time the five criterions showed similar performance where CAIC and BIC have the best performance in the case of succeed in selecting the right multivariate regression model that has the right covariance structure and in the same time the right multivariate model structure.

Key words: Multivariate regression; Information Criteria; Restricted Maximum Likelihood (REML).

INTRODUCTION

In many ways, multivariate regression is similar to MANOVA. The multivariate linear regression model composed of multiple correlated responses (dependent variables for each subject (observation)), in addition to a set of predictor variables. Multivariate linear regression allows researchers to fit a single model for each response, taking into account the correlation among the multiple responses on a given subject. The basic assumptions of multivariate regression are multivariate normality of the residuals, homogenous variances of residuals conditional on predictors, common covariance structure across observations, and independent observations. When these assumptions are satisfied, the coefficients will be unbiased, the least-squares estimates will have minimum variance, and the relationships among the coefficients will reflect the relationships among the predictors. When we deal with the multivariate linear regression model a companion to the estimation problem is the model selection problem, which consists of choosing an appropriate model from a class of candidate models to characterize the data under study. The covariance structures of the observed multiple responses makes multivariate linear regression data analysis different from univariate multiple linear regression data in term of the prediction of an individual response component given some or all of the remaining components[17].

Although the MIXED procedure of the SAS System is already widely used tools for fitting mixed effects and repeated measures models, it is also a very useful tool for fitting multivariate regression. The most advantages of using the MIXED procedure instead of stander multivariate procedure are MIXED uses observations have in complete responses, Mixed has the ability to deal with non-stander (e.g., multiple design) multivariate models, and MIXED enables researchers to fit correlated error model with different covariance structure. The MIXED procedure of the SAS System has different selections for modeling the covariance structure. The MIXED procedure of the SAS System can be used to develop either maximum likelihood (ML) or restricted maximum likelihood (REML) estimates in order to complete the analysis of the multivariate regression, where REML estimation are generally preferred to ML[25]. A lot of effort is usually needed to decide what the suitable covariance structure of the data is at the beginning of the statistical analysis. Statisticians often use information criteria such as AIC[1], BIC[22], CAIC[3], HQIC[9] to guide the selection of the covariance structure in mixed models[14,16,24]. Many studies have investigated performance of those information criteria in selection of the covariance structure considering repeated measures models[5,7,8,14]. One study compared the following, AIC, BIC, AICC, and RIC to select stander multivariate regression model when the covariance

Corresponding Author: Ali Hussein AL-Marshadi, Department of Statistics, King Abdul Aziz University, Jeddah, Saudi Arabia
Email: almarshadi-a@hotmail.com

structure is fixed[19]. Another paper by abd-Krim Seghouane[23] had developed and compared a new small sample model selection criterion for multivariate regression models to other known criterion when the covariance structure is fixed.

Our research objective is evaluating six model selection criteria to guide the selection of the best multivariate regression model that has the right covariance structure and in the same time has the right multivariate model structure.

In general form, the mixed effects linear model can be written as[11,18]:

$$Y = X\beta + ZU + e \tag{1.1}$$

where

$\beta = p \times 1$ vector of fixed effects.

$U = q \times 1$ vector of random effects.

$e = n \times 1$ vector of residuals.

$X = n \times p$ design matrix for fixed effects.

$Z = n \times q$ design matrix for random effects.

$U \sim N(0,G)$,$e \sim N(0,R)$,
$Y \sim N(X\beta,V)$, and $V=ZCZ'+R$.

When V is known, the best linear unbiased estimators (BLUE) of estimable functions $h\beta$ of the fixed effects in (1.1) are given by

$$\grave{h}\hat{\beta} = \grave{h}(X\grave{V}^{-1}X)^{-}X\grave{V}^{-1}Y, \tag{1.2}$$

with $\operatorname{var}(\grave{h}\hat{\beta}) = \grave{h}(X\grave{V}^{-1}X)^{-}h.$ (1.3)

In most applications V is unknown. Therefore, it is estimated from the data where estimators based on (1.2) are not generally BLUE[10]. Various procedures were proposed for testing hypotheses on fixed effects in mixed models with unknown V, most of which assume that V is estimated by the REML method[4,6,15]. The resulting estimates of fixed effects are often referred to as empirical BLUE (eBLUE)[10]. Standard error estimates based on (1.3) are biased downwards when V replaced by its estimate[13]. Fixed effects are estimated based on (1.2), with V replaced by a plug-in REML estimate. Null hypotheses of the form $H_0 : \grave{h}\beta = 0$ are tested by

$$F = \frac{\hat{\beta}\grave{}h[\grave{h}(X\hat{V}^{-1}X)^{-}h]^{-1}\grave{h}\hat{\beta}}{rank(h)} \sim F_{(rank(h),\upsilon)} \quad 1.3)$$

when $rank\ (h) > 1$ In general, the test statistics in (1.4) only have approximate F-distribution. The approximate denominator degree of freedom υ of F-distribution can be determined using one of the four different methods implemented in MIXED procedure of SAS. The four methods of the approximations are residual method, containment method (this is the default in MIXED), extended Satterthwaite[21] method of Giesbrecht and Burns[6] and Fai and Cornelius[4], and Kenward-Roger method[15]. Kenward and Roger[15] found good performance of their method across a number of designs. Also, Guerin and Stroup[8] recommended using the Kenward-Roger method as standard operating procedure. Therefore, Kenward-Roger method was considered in this paper for approximating the denominator degrees of freedom.

MATERIAL AND METHODS

The following model reflects the standard multivariate linear model:

$$Y = X\beta + e \tag{2.1}$$

where

$Y = n \times r$ matrix of r response variables measured on n subjects.
$X = n \times p$ design matrix for explanatory variables.
$\beta = p \times r$ matrix of regression coefficients.
$e = n \times r$ matrix of residuals whose rows are iid normal, i.e. the rows of $e \sim N(0,\Sigma)$.

In the standard multivariate linear regression model the response distribution is $Y \sim N(X\beta, I_n \otimes \Sigma)$, where the parameter space consists of the coefficient matrix β of order $p \times r$ plus the covariance matrix, $\Sigma \in PD_r$, the cone of symmetric positive semi-definite r-matrices. This multivariate regression setting is called the standard multivariate regression model because both components of the parameter space are unconstrained. In the following we give an example showing the format of the standard multivariate linear model and its relationship to the MIXED format. The example considers the format with two response variables and one explanatory variable in addition to an intercept term for three subjects.

$$\begin{bmatrix} y1_1 & y2_1 \\ y1_2 & y2_2 \\ y1_3 & y2_3 \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{bmatrix} \begin{bmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \end{bmatrix} + \begin{bmatrix} e1_1 & e2_1 \\ e1_2 & e2_2 \\ e1_3 & e2_3 \end{bmatrix}$$

To use the MIXED format, we need to write *Y,β,* and *e* as vectors and rearrange *X* accordingly as follow.

$$\begin{bmatrix} y1_1 \\ y2_1 \\ y1_2 \\ y2_2 \\ y1_3 \\ y2_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & x_1 & 0 \\ 0 & 1 & 0 & x_1 \\ 1 & 0 & x_2 & 0 \\ 0 & 1 & 0 & x_2 \\ 1 & 0 & x_3 & 0 \\ 0 & 1 & 0 & x_3 \end{bmatrix} \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{11} \\ \beta_{12} \end{bmatrix} + \begin{bmatrix} e1_1 \\ e2_1 \\ e1_2 \\ e2_2 \\ e1_3 \\ e2_3 \end{bmatrix}$$

The MIXED format in matrix notation is $\tilde{Y} = \tilde{X}\tilde{\beta} + \tilde{e}$, which is a special case of the mixed model (1.1)[25]. The situation of interest in this paper is one in which the responses are correlated in particular ways. Different covariance matrix structures of Σ were used to simulate correlated error models for the simulated study data. The following experimental design was consider because of its practical relevance. The design of the simulated experiment is described below:

There are seven correlated response variables ($y_1, y_2, y_3, y_4, y_5, y_6$, and y_7) which are related to two predictor variables (x_1 and x_2) with five different multivariate model structures and different covariance structures for the seven correlated response variables. The multivariate model structures of the simulated experiment are described as follow:

The first multivariate model structure is a standard multivariate model structure which fits seven intercepts (one for level of responses), seven slopes for x_1, and seven slops for x_2 (plus the elements of the covariance matrix of the multiple responses) as follow.

$$y_1 = \beta_{01} + \beta_{11}x_1 + \beta_{21}x_2 + e$$
$$y_2 = \beta_{02} + \beta_{12}x_1 + \beta_{22}x_2 + e$$
$$y_3 = \beta_{03} + \beta_{13}x_1 + \beta_{23}x_2 + e$$
$$y_4 = \beta_{04} + \beta_{14}x_1 + \beta_{24}x_2 + e$$
$$y_5 = \beta_{05} + \beta_{15}x_1 + \beta_{25}x_2 + e$$
$$y_6 = \beta_{06} + \beta_{16}x_1 + \beta_{26}x_2 + e$$
$$y_7 = \beta_{07} + \beta_{17}x_1 + \beta_{27}x_2 + e$$

The second multivariate model structure is a standard multivariate model structure which fits seven intercepts (one for level of responses), and seven slopes for x_1 (plus the elements of the covariance matrix of the multiple responses) as follow.

$$y_1 = \beta_{01} + \beta_{11}x_1 + e$$
$$y_2 = \beta_{02} + \beta_{12}x_1 + e$$
$$y_3 = \beta_{03} + \beta_{13}x_1 + e$$
$$y_4 = \beta_{04} + \beta_{14}x_1 + e$$
$$y_5 = \beta_{05} + \beta_{15}x_1 + e$$
$$y_6 = \beta_{06} + \beta_{16}x_1 + e$$
$$y_7 = \beta_{07} + \beta_{17}x_1 + e$$

The third multivariate model structure is a standard multivariate model structure which fits seven intercepts (one for level of responses), and seven slops for x_2 (plus the elements of the covariance matrix of the multiple responses) as follow.

$$y_1 = \beta_{01} + \beta_{11}x_2 + e$$
$$y_2 = \beta_{02} + \beta_{12}x_2 + e$$
$$y_3 = \beta_{03} + \beta_{13}x_2 + e$$
$$y_4 = \beta_{04} + \beta_{14}x_2 + e$$
$$y_5 = \beta_{05} + \beta_{15}x_2 + e$$
$$y_6 = \beta_{06} + \beta_{16}x_2 + e$$
$$y_7 = \beta_{07} + \beta_{17}x_2 + e$$

The forth multivariate model structure is a non-standard multivariate model structure which is called "multiple design". It allows each response variable to have a different set of explanatory variables as follow.

$$y_1 = \beta_{01} + e$$
$$y_2 = \beta_{02} + \beta_{12}x_1 + e$$
$$y_3 = \beta_{03} + \beta_{13}x_2 + e$$
$$y_4 = \beta_{04} + \beta_{14}x_1 + \beta_{24}x_2 + e$$
$$y_5 = \beta_{05} + \beta_{15}x_1 + \beta_{25}x_2 + e$$
$$y_6 = \beta_{06} + \beta_{16}x_1 + \beta_{26}x_2 + e$$
$$y_7 = \beta_{07} + \beta_{17}x_1 + \beta_{27}x_2 + e$$

The fifth multivariate model structure is also a non-standard multivariate model structure which is called "multiple design". It is also called "multiple design" which allows each response variable to have a different set of explanatory variables as follow.

$$y_1 = \beta_{01} + \beta_{11}x_2 + e$$
$$y_2 = \beta_{02} + \beta_{12}x_1 + e$$
$$y_3 = \beta_{03} + \beta_{13}x_1 + \beta_{23}x_2 + e$$
$$y_4 = \beta_{04} + \beta_{14}x_1 + e$$
$$y_5 = \beta_{05} + \beta_{15}x_1 + \beta_{25}x_2 + e$$
$$y_6 = \beta_{16}x_1 + \beta_{26}x_2 + e$$
$$y_7 = \beta_{07} + e$$

The multivariate regression analyses for the first multivariate model structure design can be implemented by the following example SAS code[25]:

```
PROC MIXED DATA = one;
CLASS time;
MODEL y = time time* x1 time*x2 / noint notest ddfm = kr;
REPEATED time / type = UN subject = subject;
```

The multivariate regression analyses for the second multivariate model structure design can be implemented by the following example SAS code[25]:

```
PROC MIXED DATA = one;
CLASS time;

MODEL y = time time*x1 / noint notest ddfm = kr;
REPEATED time / type = UN subject = subject;
```

The multivariate regression analyses for the third multivariate model structure design can be implemented by the following example SAS code[25]:

```
PROC MIXED DATA = one;
CLASS time;
MODEL y = time time*x2 / noint notest ddfm = kr;
REPEATED time / type = UN subject = subject;
```

Note: The class variable "time" in the first, second and third structure is used to identify the multiple responses.

The multivariate regression analyses for the forth multivariate model structure design can be implemented by the following example SAS code[25]:

```
PROC MIXED DATA = one;
CLASS time;
MODEL y = time1 time2 time2*x1 time3 time3*x2 time4 time4*x1 time4*x2 time5 time5*x1 time5*x2 time6 time6*x1 time6*x2 time7 time7*x1 time7*x2 / noint notest ddfm = kr;
REPEATED time / type = UN subject = subject;
```

The multivariate regression analyses for the fifth multivariate model structure design can be implemented by the following example SAS code[25]:

```
PROC MIXED DATA = one;
CLASS time;
MODEL y = time1 time1*x2 time2 time2*x1 time3 time3*x1 time3*x2 time4 time4*x1 time5 time5*x1 time5*x2 time6*x1 time6*x2 time7 / noint notest ddfm = kr;
REPEATED time / type = UN subject = subject;
```

Note: The variable "time" is replaced in the forth and fifth structure by individual 0-1 dummy variables, one for each responses variable.

In MIXED procedure, users find five model selection criteria available, which give users tools can be used to select an appropriate model. The five model selection criteria are:

1. Akaike[1] Information Criterion (AIC)
2. Schwarz[22] Bayesian Information Criterion (BIC)
3. Bozdogan[3] Corrected Akaike Information Criterion (CAIC)
4. Hannan and Quinn[9] Information Citerion (HQIC) and
5. Hurvich and Tsai[12] The Corrected Akaike information Criterion (AICC).

Our study concerns with comparing the five information criterions in terms of their ability to identify the right multivariate regression model that has the right covariance structure and in the same time the right multivariate model structure from both standard and non-standard multivariate model structures i.e. from multivariate model structures with both "single design" such as the first three structures and "multiple design" such as the last two structures.

3. The Simulation Study: A simulation study of PROC MIXED's mixed model analysis of multivariate regression data was conducted to compare the five model selection criteria in terms of their percentage of number of times that they identify the right covariance structure and in the same time the right multivariate model structure.

Correlated multivariate normal data were generated according to MIXED format model. There were 35 scenarios to generate data involving five multivariate regression model structures and seven covariance structures with one setting of covariance matrix parameter values for each covariance structure and sample sizes 40 (n =40 subjects). The covariance structures were Independent Errors (VC), Compound Symmetry (CS), Heterogeneous Compound Symmetry (CSH), First-Order Autoregressive (AR(1)), Heterogeneous First-Order Autoregressive (ARH(1)), , Banded Main Diagonal (UN(1)) and Unstructured (UN). The 7 settings of covariance matrix parameter values are given in Table 1. For each scenario, we simulated 3000 datasets. SAS[20] code was written to generate the datasets according to the described design. We will consider the case when we have 12 subjects as an example to explain the process of generating the datasets. A 12 7x1 vector of standard normal random deviates were generated using SAS's NORMAL function. Denoted the vector:

$$\varepsilon_i = \left[\varepsilon_{1i}\,\varepsilon_{2i}\,\varepsilon_{3i}\,\varepsilon_{4i}\,\varepsilon_{5i}\,\varepsilon_{6i}\,\varepsilon_{7i}\right]'$$

where i = 1,2,3,...,12. Note that the 12 represents the 12 subjects and the 7 represents the 7 levels of time effect within each subject. Then the 12 7x1 vectors of residuals for model (2.1) were calculated as

$e_i = \Sigma^{\frac{1}{2}}\varepsilon_i$; $i = 1,2,3,\ldots,12$, where:

$\Sigma^{\frac{1}{2}}$ is the Cholevsky decomposition of Σ, and

Σ is the covariance matrix of multiple response variable "time".

Therefore, the vector e_i is defined as the rows of the residuals matrix, e, such that $e \sim N(0,\Sigma)$. The fixed portion of the model, $X\beta$, is added to the residuals matrix, e, according to the model structure to give the vector of response, Y. The first explanatory variable was as indicator variable with two levels and the second explanatory variable was randomly generated from normal distribution with mean equal 30 and variance equal to 5. Each one of the 3000 generated data sets was fitted to all the possible combination of the selected model structures and covariance structures for the situation mentioned before. Then each one of the five information criteria was calculated in order to compare the performance of the five information criteria in identify the right multivariate regression model that has the right covariance structure and in the same time the right multivariate model structure.

The 7 settings of the covariance matrix are given in Table 1 which can be categorized to seven covariance structures. The first one, (Setting No. 1) represents Compound Symmetry (CS) covariance structure with different choices of parameters. The second one, (Setting No. 2) represents First-Order Autoregressive (AR(1)) covariance structure with different choices of parameters. The third one, (Setting No. 3) represents Heterogeneous First-Order Autoregressive (ARH(1)) covariance structure with different choices of parameters. The fourth one, (Setting No. 4) represents Heterogeneous Compound Symmetry (CSH) covariance structure with different choices of parameters. The fifth one, (Setting No. 5) represents Independent Errors (VC) covariance structure. The sixth one, (Setting No. 6) represents Banded Main Diagonal (UN(1)) covariance structure. The seventh one, (Setting No. 7) represents Unstructured (UN) covariance structure.

RESULTS AND DISCUSSIONS

Results: Table 2-37 present the percentages of models that were chosen by the five criteria for the data generated by each model considered in the study. Tables 2-37 indicate that although the percentages of identifying the right multivariate regression model that has the right covariance structure and in the same time the right multivariate model structure from both standard and non-standard multivariate model structures were low overall with the five criterions, great percentages upward were observed with specific models that involve the indicator variable. In the same time all the five criterions showed similar performance for these specific models where CAIC and BIC have the best performance overall. This result is not surprising since the challenge in working with multivariate regression model comes from the fast increasing in the count of the parameters as the complexity of the model increases comparing to the univariate regression model. This fast increase causes many model selection criteria to perform poorly[2].

5. Conclusion: In our simulation, we considered multivariate regression, looking at the performance of the five information criteria in identify the right multivariate regression model that has the right

Table 1: The setting of seven Covariance Matrix structures used in the Simulations

Setting No.	Covariance Matrix	Setting No.	Covariance Matrix
1.	$\begin{bmatrix} 16 & 12.8 & 12.8 & 12.8 & 12.8 & 12.8 & 12.8 \\ 12.8 & 16 & 12.8 & 12.8 & 12.8 & 12.8 & 12.8 \\ 12.8 & 12.8 & 16 & 12.8 & 12.8 & 12.8 & 12.8 \\ 12.8 & 12.8 & 12.8 & 16 & 12.8 & 12.8 & 12.8 \\ 12.8 & 12.8 & 12.8 & 12.8 & 16 & 12.8 & 12.8 \\ 12.8 & 12.8 & 12.8 & 12.8 & 12.8 & 16 & 12.8 \\ 12.8 & 12.8 & 12.8 & 12.8 & 12.8 & 12.8 & 16 \end{bmatrix}$	2.	$\begin{bmatrix} 16 & 12.8 & 10.24 & 8.192 & 6.5536 & 5.24288 & 4.194304 \\ 12.8 & 16 & 12.8 & 10.24 & 8.192 & 6.5536 & 5.24288 \\ 10.24 & 12.8 & 16 & 12.8 & 10.24 & 8.192 & 6.5536 \\ 8.192 & 10.24 & 12.8 & 16 & 12.8 & 10.24 & 8.192 \\ 6.5536 & 8.192 & 10.24 & 12.8 & 16 & 12.8 & 10.24 \\ 5.24288 & 6.5536 & 8.192 & 10.24 & 12.8 & 16 & 12.8 \\ 4.194304 & 5.24288 & 6.5536 & 8.192 & 10.24 & 12.8 & 16 \end{bmatrix}$
3.	$\begin{bmatrix} 4 & 4.8 & 5.12 & 5.12 & 4.9152 & 4.58752 & 4.194304 \\ 4.8 & 9 & 9.6 & 9.6 & 9.216 & 8.6016 & 7.86432 \\ 5.12 & 9.6 & 16 & 16 & 15.36 & 14.336 & 13.1072 \\ 5.12 & 9.6 & 16 & 25 & 24 & 22.4 & 20.48 \\ 4.9152 & 9.216 & 15.36 & 24 & 36 & 33.6 & 30.72 \\ 4.58752 & 8.6016 & 14.336 & 22.4 & 33.6 & 49 & 44.8 \\ 4.194304 & 7.86432 & 13.1072 & 20.48 & 30.72 & 44.8 & 64 \end{bmatrix}$	4.	$\begin{bmatrix} 4 & 4.8 & 6.4 & 8 & 9.6 & 11.2 & 12.8 \\ 4.8 & 9 & 9.6 & 12 & 14.4 & 16.8 & 19.2 \\ 6.4 & 9.6 & 16 & 16 & 19.2 & 22.4 & 25.6 \\ 8 & 12 & 16 & 25 & 24 & 28 & 32 \\ 9.6 & 14.4 & 19.2 & 24 & 36 & 33.6 & 38.4 \\ 11.2 & 16.8 & 22.4 & 28 & 33.6 & 49 & 44.8 \\ 12.8 & 19.2 & 25.6 & 32 & 38.4 & 44.8 & 64 \end{bmatrix}$
5.	$\begin{bmatrix} 16 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 16 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 16 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 16 \end{bmatrix}$	6.	$\begin{bmatrix} 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 16 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 36 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 49 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 64 \end{bmatrix}$
7.	$\begin{bmatrix} 4 & 2.4 & 4.8 & 8 & 8.4 & 7 & 4.96 \\ 2.4 & 9 & 2.4 & 1.5 & 2.7 & 7.35 & 10.8 \\ 4.8 & 2.4 & 16 & 3.4 & 10.08 & 15.4 & 6.48 \\ 8 & 1.5 & 3.4 & 25 & 18.9 & 16.45 & 9.2 \\ 8.4 & 2.7 & 10.08 & 18.9 & 36 & 4.62 & 22.56 \\ 7 & 7.35 & 15.4 & 16.45 & 4.62 & 49 & 16.24 \\ 4.96 & 10.8 & 6.48 & 9.2 & 22.56 & 16.24 & 64 \end{bmatrix}$		

Table 2: Percent of each Models Chosen by Five Criteria for the Data of model (1,1).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	93.22	99.89	99.89	99.22	94.89
1	4	6.22	0.11	0.11	0.78	5.00
1	7	0.56	0	0	0	0.11

Table 3: Percent of each Models Chosen by Five Criteria for the Data of model (2,1).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	4	0.11	0	0	0	0.11
2	1	92.77	100	100	98.78	94.55
2	4	6.45	0	0	1.22	5.12
2	7	0.67	0	0	0	0.22

Table 4: Percent of each Models Chosen by Five Criteria for the Data of model (3,1).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	93.11	99.78	99.78	99.11	94.78
3	1	0	0	0	0	0
1	4	6.22	0.11	0.11	0.78	5.00
1	7	0.56	0	0	0	0.11
4	2	0.11	0.11	0.11	0.11	0.11

Table 5: Percent of each Models Chosen by Five Criteria for the Data of model (4,1).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	73.56	78.56	78.56	78.00	74.33
1	4	5.00	0.11	0.11	0.67	4.00
1	7	0.11	0	0	0	0
4	1	19.56	21.33	21.33	21.11	20.44
4	4	1.33	0	0	0.22	1.00
4	7	0.44	0	0	0	0.22

Table 6: Percent of each Models Chosen by Five Criteria for the Data of model (5,1).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	86.56	92.56	92.56	92.00	88.00
1	4	5.44	0.11	0.11	0.67	4.56
5	1	0	0	0	0	0
1	7	0.56	0	0	0	0.11
5	1	6.67	7.33	7.33	7.33	6.78
5	4	0.78	0	0	0	0.56

Table 8: Percent of each Models Chosen by Five Criteria for the Data of model (1,2).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	2	92.56	99.78	100	98.44	94.00
1	3	6.89	0.22	0	1.56	5.67
1	7	0.56	0	0	0	0.33

Table 9: Percent of each Models Chosen by Five Criteria for the Data of model (2,2).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	2	0.11	0.11	0.11	0.11	0.11
2	2	92.56	99.78	99.89	98.11	94.11
2	3	6.78	0.11	0	1.78	5.56
2	7	0.56	0	0	0	0.22

Table 10: Percent of each Models Chosen by Five Criteria for the Data of model (3,2).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	2	92.56	99.78	100	98.44	94.00
1	3	6.89	0.22	0	1.56	5.67
1	7	0.56	0	0	0	0.33
3	2	0	0	0	0	0

Table 11: Percent of each Models Chosen by Five Criteria for the Data of model (4,2).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	2	79.98	86.21	86.43	84.98	80.87
4	2	12.57	13.57	13.57	13.46	13.01
1	3	5.90	0.22	0	1.56	5.01
4	3	1.00	0	0	0	0.78
1	7	0.44	0	0	0	0.22
4	7	0.11	0	0	0	0.11

Table 12: Percent of each Models Chosen by Five Criteria for the Data of model (5,2).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	0.11	0	0	0	0.11
1	2	92.44	99.78	100	98.44	93.89
1	3	6.89	0.22	100	1.56	5.67
1	7	0.56	0	0	0	0.33
5	2	0	0	0	0	0

Table 13: Percent of each Models Chosen by Five Criteria for the Data of model (1,3).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	2	0	0.11	0.11	0	0
1	3	97.89	99.89	99.89	99.78	99.56
1	4	0.11	0	0	0.11	0.11
1	7	2.00	0	0	0.11	0.33

Table 14: Percent of each Models Chosen by Five Criteria for the Data of model (2,3).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	3	0.44	0.44	0.44	0.44	0.44
2	3	97.44	99.56	100	99.44	99.00
2	7	2.11	0	0	0.11	0.56

Table 15: Percent of each Models Chosen by Five Criteria for the Data of model (3,3).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	3	97.89	99.89	99.89	99.78	99.56
3	3	0.11	0.11	0.11	0.11	0.11
1	7	2.00	0	0	0.11	0.33

Table 16: Percent of each Models Chosen by Five Criteria for the Data of model (4,3).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	3	32.56	33.11	33.11	33.11	33.11
4	3	65.44	66.78	66.78	66.67	66.44
4	4	0.11	0	0	0.11	0.11
1	7	0.33	0	0	0	0
4	7	1.56	0	0	0.11	0.33

Table 17: Percent of each Models Chosen by Five Criteria for the Data of model (5,3).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	0.11	0.11	0.11	0.11	0.11
1	3	97.89	99.89	99.89	99.78	99.56
1	7	2.00	0	0	0.11	0.33
5	3	0	0	0	0	0

Table 18: Percent of each Models Chosen by Five Criteria for the Data of model (1,4).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	4	98.11	100	100	100	99.56
1	7	1.89	0	0	0	0.44

Table 19: Percent of each Models Chosen by Five Criteria for the Data of model (2,4).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	4	0.22	0.22	0.22	0.22	0.22
2	4	98.11	99.78	99.78	99.78	99.33
2	7	1.67	0	0	0	0.44

Table 20: Percent of each Models Chosen by Five Criteria for the Data of model (3,4).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	4	98.11	100	100	100	99.56
1	7	1.89	0	0	0	0.44
3	4	0	0	0	0	0

Table 21: Percent of each Models Chosen by Five Criteria for the Data of model (4,4).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	4	29.00	29.33	29.33	29.33	29.22
4	4	69.11	70.67	70.67	70.67	70.33
1	7	0.33	0	0	0	0
4	7	1.56	0	0	0	0.44

Table 22: Percent of each Models Chosen by Five Criteria for the Data of model (5,4).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	0.11	0	0	0.11	0.11
1	4	97.89	99.89	99.89	99.78	99.33
5	4	0.11	0.11	0.11	0.11	0.11
1	7	1.89	0	0	0	0.44

Table 23: Percent of each Models Chosen by Five Criteria for the Data of model (1,5).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	11.89	4.22	2.67	7.89	11.67
1	2	12.56	5.67	3.00	9.33	12.44
1	3	0.89	0	0	0.22	0.89
1	4	0.56	0	0	0	0.56
1	5	69.22	90.00	94.33	81.56	70.67
1	6	4.44	0.11	0	1.00	3.56
1	7	0.44	0	0	0	0.22

Table 24: Percent of each Models Chosen by Five Criteria for the Data of model (2,5).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	0.11	0.11	0.11	0.11	0.11
2	1	11.35	4.78	3.00	7.12	11.01
1	2	0.44	0.33	0.11	0.33	0.44
2	2	11.79	5.23	2.34	9.12	11.68
1	3	0.11	0	0	0.11	0.11
2	3	1.11	0	0	0	0.89
2	4	0.44	0	0	0.22	0.44
1	5	1.89	2.45	2.67	2.22	2.00
2	5	67.96	86.87	91.77	79.42	69.19
1	6	0.33	0	0	0.11	0.22
2	6	4.00	0.22	0	1.22	3.78
2	7	0.44	0	0	0	0.11

Table 25: Percent of each Models Chosen by Five Criteria for the Data of model (3,5).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	11.89	4.22	2.67	7.89	11.67
1	2	12.56	5.67	3.00	7.89	12.44
1	3	0.89	0	0	0.22	0.89
1	4	0.56	0	0	0	0.56
1	5	69.22	90.00	94.33	81.56	70.67
1	6	4.44	0.11	0	1.00	3.56
1	7	0.44	0	0	0	0.22
3	5	0	0	0	0	0

Table 26: Percent of each Models Chosen by Five Criteria for the Data of model (4,5).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	11.78	4.22	2.67	7.89	11.67
1	2	12.44	5.67	3.00	9.33	12.33
4	2	0.11	0	0	0	0.11
1	3	0.89	0	0	0.22	0.89
4	3	0.11	0	0	0	0
1	4	0.56	0	0	0	0.56
1	5	69.11	89.78	94.11	81.33	70.56
4	5	0.11	0.22	0.22	0.22	0.11
1	6	4.44	0.11	0	1.00	3.56
1	7	0.44	0	0	0	0.11

Table 27: Percent of each Models Chosen by Five Criteria for the Data of model (5,5).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	11.89	4.22	2.67	7.89	11.67
1	2	12.56	5.67	3.00	9.33	12.44
1	3	0.89	0	0	0.22	0.89
1	4	0.56	0	0	0	0.56
1	5	69.22	90.00	94.33	81.56	70.67
1	6	4.44	0.11	0	1.00	3.56
1	7	0.44	0	0	0	0.22
5	5	0	0	0	0	0

Table 28: Percent of each Models Chosen by Five Criteria for the Data of model (1,6).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	3	14.79	6.01	3.34	10.57	13.35
1	4	11.68	4.56	3.11	8.23	11.01
1	6	72.41	89.43	93.55	81.09	75.42
1	7	1.11	0	0	0.11	0.22

Table 29: Percent of each Models Chosen by Five Criteria for the Data of model (2,6).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	3	1.22	0.44	0.22	0.78	1.11
2	3	13.00	5.89	2.44	9.56	12.56
1	4	0.67	0.22	0.22	0.56	0.67
2	4	10.56	4.22	2.89	7.89	9.78
1	6	5.22	6.44	6.67	5.89	5.44
2	6	67.89	82.78	87.56	75.22	70.22
2	7	1.44	0	0	0.11	0.22

Table 30: Percent of each Models Chosen by Five Criteria for the Data of model (3,6).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	3	14.78	6.00	3.33	10.56	13.33
1	4	11.67	4.56	3.11	8.22	11.00
1	6	72.44	89.44	93.56	81.11	75.44
1	7	1.11	0	0	0.11	0.22
3	6	0	0	0	0	0

Table 31: Percent of each Models Chosen by Five Criteria for the Data of model (4,6).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	3	7.89	3.89	2.11	6.22	7.33
4	3	6.67	2.11	1.33	4.33	5.89
1	4	6.44	2.33	1.56	4.33	6.00
4	4	5.22	2.33	1.56	4.00	5.00
1	6	40.56	48.89	51.56	44.56	41.78
4	6	32.11	40.44	41.89	36.33	33.78
1	7	0.33	0	0	0	0
4	7	0.78	0	0	0.22	0.22

Table 32: Percent of each Models Chosen by Five Criteria for the Data of model (5,6).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	0.11	0	0	0	0
1	3	14.78	6.00	3.33	10.56	13.33
1	4	11.67	4.56	3.11	8.22	11.00
1	6	72.33	89.44	93.56	81.11	75.44
1	7	1.11	0	0	0.11	0.22
5	6	0	0	0	0	0

Table 33: Percent of each Models Chosen by Five Criteria for the Data of model (1,7).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	7	100	100	100	100	100

Table 34: Percent of each Models Chosen by Five Criteria for the Data of model (2,7).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	0.11	0.11	0.11	0.11	0.11
1	7	0.44	0.44	0.44	0.44	0.44
2	7	99.44	99.44	99.44	99.44	99.44

Table 35: Percent of each Models Chosen by Five Criteria for the Data of model (3,7).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	7	99.89	99.89	99.89	99.89	99.89
3	7	0.11	0.11	0.11	0.11	0.11

Table 36: Percent of each Models Chosen by Five Criteria for the Data of model (4,7).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	1	0.11	0.11	0.11	0.11	0.11
1	7	56.22	56.22	56.22	56.22	55.33
4	7	43.67	43.67	43.67	43.67	44.56

Table 37: Percent of each Models Chosen by Five Criteria for the Data of model (5,7).

Model Analysis		Criterion				
Model Structure	Covariance Structure	AIC	BIC	CAIC	HQIC	AICC
1	7	100	100	100	100	100
5	7	0	0	0	0	0

covariance structure and in the same time the right multivariate model structure from both standard and non-standard multivariate model structures. The main result of our article is that overall the percentages of identifying the right multivariate regression model that has the right covariance structure and in the same time the right multivariate model structure from both standard and non-standard multivariate model structures were low except for specific models that involve the indicator variable. In the same time the five criterions showed similar performance where CAIC and BIC have the best performance overall in the case of succeed in selecting the right multivariate regression model that has the right covariance structure and in the same time the right multivariate model structure. Hence, finding a selection criterion to selects the right multivariate regression model that has jointly the right covariance structure and the right multivariate model structure from both standard and non-standard multivariate model structures would be a potential research area for future study.

REFERENCES

1. Akaike, H., 1974. A new look at the statistical model identification. IEEE Transaction on Automatic Control, AC-19: 716 - 723.
2. Bedrik, E.J. and C.L. Tsai, 1994. Model selection for multivariate regression in small samples, Biometrics, 50: 226-231.
3. Bozdogan, H., 1987. Model selection and Akaike's information criterion (aic): the general theory and its analytical extensions, Psychometrika, 52: 345 - 370.
4. Fai, A.H.T. and P.L. Cornelius, 1996. Approximate F-Tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-plot Experiments. J. Statist. Compu. and Simu., 54: 363-378.
5. Ferron, J., R. Dailey, Q. Yi, 2002. Effects of misspecifying the first-level error structure in two-level models of change. Multivariate Behav. Res., 37: 379-403.
6. Giesbrecht, F.G. and J.C. Burns, 1985. Two stage Analysis Based on a Mixed Model: Large Sample Asymptotic Theory and Small-Sample Simulation Results. Biometrics, 41: 477-486.
7. Gomez, E.V., G.B. Schaalje and G.W. Fellingham, 2005. Performance of the Kenward-Roger method when the covariance structure is selected using AIC and BIC. Commun. Statist. Simul. Computat., 34: 377-392.
8. Guerin, L. and W.W. Stroup, 2000. A simulation study to evaluate PROC MIXED Analysis of repeated measures data. In Proceedings of the 2000 Conference on Applied Statistics in Agriculture, pp: 170-203. Manhattan, KS.: Kansas State University.
9. Hannan, E.J. and B.G. Quinn, 1979. The Determination of the order of an autoregression, J. Royal Statist. Soc., Ser., B, 41: 190 - 195.
10. Henderson, C.R., 1963. Selection Index and Expected Genetic Advance. Statistical Gentics and Plant Breeding, NAS-NRC Publ. No. 982: 141-163.
11. Henderson, C.R., 1990. Statistical Method in Animal Improvement: Historical Overview. In Advances in Statistical Methods for Genetic Improvement of Livesstock, New York: Springer.
12. Hurvich, C.M. and C.L. Tsai, 1989. Regression and time series model selection in small samples. Biometrika, 76: 297-307.
13. Kacker, R.N. and D.A. Harville, 1984. Approximations for standard errors of estimators of fixed and random effects in mixed linear models. J. Amer. Statist. Assoc., 79: 853-862.
14. Keselman, H.J., J. Algina, R.K. Kowalchuk, R.D. Wolfinger, 1999. The analysis of repeated measurements: a comparison of mixed-model Satterthwaite F test and a nonpooled adjusted degrees of freedom multivariate test. Commun. Statist. Theor. Meth., 28: 2967-2999.
15. Kenward, M.G. and J.H. Rogers, 1997. Small sample inference for fixed effect from restricted maximum likelihood. Biometrics, 53: 983-97.
16. Littell, R.C., J. Pendergast, R. Natarajan, 2000. Modeling covariance structure in the analysis of repeated measures data. Statist. Med., 19: 1793-1819.
17. McCullagh, Peter, 2006. Structured covariance matrices in multivariate regression models. Technical reports online: http://www.stat.uchicago.edu/~pmcc/reports/.
18. McCulloch, C. and S. Searle, 2001. Generalized, Linear, and Mixed Models. Wiley, New York.
19. Rahman Azari, Lexin Li and Chih-Ling Tsai 2006. Longitudinal data model selection. Computational Statistics & Data Analysis, 50: 3053-3066.
20. SAS Institute, 2003. *SAS/STAT, Release 9.1*. Cary NC: SAS Institute Inc.
21. Satterthwaite, F.E., 1941. Synthesis of Variance. Psychometrika, 6: 353-367.
22. Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics, 6: 461 - 464.
23. Seghouane Abd-Krim, 2006. Multivariate regression model selection from small samples using kullback's symmetric divergence, Signal Processing, 86: 2074-2084.
24. Singer, J.D., 1998. Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. J. Educat. Behavi. Statist., 24: 323-325.
25. Wright, S. Paul, 1998. Multivariate analysis using the mixed procedure. In Proceedings of the SUGI-23 Conference in statistics, data analysis, and modeling, pp: 1-5. Nashville, Tennessee.